ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12239

A genome-wide association study for the incidence of persistent bovine viral diarrhea virus infection in cattle[1]

E. Casas*, B. E. Hessman†, J. W. Keele‡ and J. F. Ridpath*

*National Animal Disease Center, ARS, USDA, Ames, IA, USA. †Haskell County Animal Hospital, Sublette, KS, USA. ‡US Meat Animal Research Center, ARS, USDA, Clay Center, NE, USA.

Summary

Bovine viral diarrhea viruses (BVDV) comprise a diverse group of viruses that cause disease in cattle. BVDV may establish both transient and persistent infections depending on the developmental stage of the animal at exposure. The objective was to determine whether genomic regions harboring single nucleotide polymorphisms (SNPs) could be associated with the presence or absence of persistent BVDV infection. A genome-wide association approach based on 777 000 SNP markers was used. Samples of animals identified as positive ($n = 1200$) or negative ($n = 1200$) for the presence of BVDV in skin samples ($n = 1200$) were used. DNA samples were combined in 24 pools (100 animals per pool). One SNP, significant at the 5 percent genome-wide level ($P = 9.41 \times 10^{-8}$), was detected on chromosome 14, located at position 80 675 176 bp. Fifteen SNPs, residing on chromosomes 1, 2, 6, 8, 10, 15 and 18, were moderately associated ($P < 1 \times 10^{-5}$) with persistent BVDV infection. Results show that genes harboring or neighboring significant SNPs are involved in leucopenia, signal transduction, RNA splicing and DNA methylation processes.

Keywords BVDV, diseases, genome scan, persistent infection

Introduction

Bovine viral diarrhea viruses (species BVDV1 and BVDV2; family *Flaviviridae*, genus *Pestivirus*) are an economically important problem. BVDV infections have a variety of clinical symptoms of varying severity, ranging from subclinical to highly fatal, depending on the viral strain, immune status, maturation stage and presence of secondary pathogens. Clinical presentations associated with BVDV infection can involve single or multiple organ systems. All BVDV infections result in the destruction of immune tissue. BVDV infections may be transient or persistent. Persistent infection occurs from *in utero* infection of a dam that becomes viremic from BVDV between approximately 42 and 125 days of gestation. If the pregnancy survives, the calf will be born immunotolerant to the specific virus and will be a lifelong shedder (McClurkin *et al.* 1984). Although many persistently infected (PI) animals die in the first 30 days of life, some survive and enter production cattle operations. Exposure of pen mates to PI animals has significant negative economic impact (Hessman *et al.* 2009). For this reason, cattle production units institute BVDV detection protocols aimed at finding and eliminating PI animals.

Genome-wide association studies are now possible due to the availability of technology that permits high-throughput genotyping of SNPs. This technology allows deciphering the genetics behind the expression of economically important traits. Genomic regions associated with diseases in cattle have been identified using this approach (Blaschek *et al.*, 2011; Finlay *et al.*, 2012; Minozzi *et al.*, 2012). The objective was to determine genomic regions harboring single nucleotide polymorphisms (SNPs) associated with presence or absence of persistent BVDV infections based on analysis of tissue samples collected in the course of BVDV surveillance in commercial beef production units.

Materials and methods

Animals

Animal experimental procedures were approved according to the commercial feedlot standard operating procedures. The feedlot's management approved the study prior to initiation. Cattle were acquired by a starter feedlot in

Address for correspondence

E. Casas, National Animal Disease Center, ARS, USDA, 1920 Dayton Ave., Ames, IA 50010, USA.
E-mail: Eduardo.Casas@ARS.USDA.GOV

[1]Mention of trade name, proprietary product, or specified equipment does not constitute a guarantee or warranty by the USDA and does not imply approval to the exclusion of other products that may be suitable. USDA is an Equal Opportunity Employer.

Accepted for publication 17 September 2014

southwestern Kansas, and samples were collected in conjunction with typical practices of the feedlot. Cattle originated from auction market facilities in multiple states in the southern and southeastern United States and were tested upon entry into the feedlot. All animals used were cross-bred with unknown pedigree. The average weight at the time of arrival to the feedlot was 233.2 ± 1.7 kg. A total of 21 743 ear notch samples were collected from the summer of 2004 to the summer of 2008 (Hessman *et al.* 2009).

Trait

Cattle were tested for BVDV according to the procedure described by Fulton *et al.* (2006). Ear notch samples were collected and placed in PBS solution for testing to detect BVDV antigen by use of an antigen-capture ELISA (ACE) test. A second set of samples was collected within 48 h from cattle identified as positive by ACE. These samples included a second ear notch, a portion of which was fixed in formalin and tested by immunohistochemistry and a portion of which was used for a second ACE test, and serum, which was used for virus isolation and reverse transcription PCR amplification of viral RNA (Hessman *et al.* 2009). Ear notch samples from 1200 animals identified as positive for the presence of BVDV and 1200 from animals identified as negative were sent to the National Animal Disease Center, Ames, Iowa. For the purpose of this study, calves diagnosed as positive were classified as 'affected' by BVDV, whereas those diagnosed as negative were classified as 'unaffected.'

Genomic screening

Ear notch samples collected in the course of the BVDV surveillance program were stored at −20 °C in PBS solution. The QIAGEN Dneasy Blood and Tissue kit was used to extract DNA from samples. Concentration of extracted DNA was measured using a fluorometer.

Twelve pools of DNA were generated for affected animals and 12 pools for unaffected animals. Samples were allocated randomly in pools based on presence or absence of BVDV. One hundred animals were combined in each pool. Pools were genotyped with the Illumina BovineHD chip at a commercial biotechnology facility (Geneseek). With a pool size of 100 animals per pool and an experiment with 24 pools, the statistical power achieved was 80 percent for detecting allele frequency differences of 5 percent (McDaneld *et al.* 2012).

To establish the consistency of the procedure, 384 animals (192 affected and 192 unaffected) were randomly selected from the population under study. Samples were individually genotyped for significant SNPs (Table 1) using a primer extension method with mass spectrometry-based analysis of the extension products on a MassARRAY system, as suggested by the manufacturer (Sequenom, Inc.) and as described by Stone *et al.* (2002). A universal mass tag sequence was added to the 5′ end of each gene-specific amplification primer sequence as recommended by the manufacturer. Genotypes for each animal were collected. Samples were genotyped only once, and given the results, there was no attempt to pursue additional genotyping.

Table 1 Single nucleotide polymorphism (SNP) identification, chromosome, position in the chromosome, allele frequency of *B* allele in the unaffected pools (unaffected), allele frequency of *B* allele in affected pools (affected) and significance for incidence of persistent bovine viral diarrhea virus.

SNP	BTA	Position (bases)	Allele frequency (unaffected)	Allele frequency (affected)	Significance[1]	Genes within 100 Kb on either side[2]
Rs43226513	1	32138151	0.243	0.159	3.32×10^{-6}	–
Rs136101709	1	154175502	0.197	0.281	4.92×10^{-6}	***METTL6***, *EAF1*, *COLQ*
Rs43287585	1	156065876	0.325	0.216	1.19×10^{-6}	***TBC1D5***
Rs133794462	2	105626910	0.771	0.878	6.38×10^{-6}	*TNP1*
Rs135524200	2	107608745	0.383	0.232	2.93×10^{-6}	*PRKAG3*, *WNT6*, *WNT10A*, *FEV*, *CRYBA2*, *MIR375*
Rs109923296	2	107747873	0.254	0.106	7.16×10^{-7}	*CCDC108*, *IHH*, ***NHEJ1***
Rs135783786	6	79369511	0.220	0.354	9.32×10^{-6}	*LPHN3*
Rs134088824	6	79385615	0.658	0.500	7.08×10^{-7}	*LPHN3*
Rs135088957	6	79389382	0.610	0.474	5.67×10^{-6}	*LPHN3*
Rs42255579	6	99531916	0.646	0.498	3.84×10^{-6}	*SEC31A*, *THAP9*, ***LIN54***, *MIR2447*, *COPS4*
Rs109480556	8	99789997	0.304	0.410	5.95×10^{-6}	–
Rs135138396	10	65281936	0.312	0.220	3.47×10^{-6}	*SLC30A4*, *MIR147*, ***SPATA5L1***, *GATM*
Rs136716397	14	80675176	0.853	0.735	9.41×10^{-8}	***RALYL***
Rs110434337	15	1385504	0.472	0.388	7.93×10^{-6}	–
Rs109671280	15	35368078	0.063	0.138	8.13×10^{-7}	*KCNC1*, *MYOD1*, *OTOG*
Rs110572575	18	44427437	0.246	0.161	1.35×10^{-6}	*KCTD15*

[1]Strong association threshold, $P < 5 \times 10^{-7}$; moderate association threshold, $P < 1 \times 10^{-5}$.
[2]Genes in bold indicate the SNP was located within the gene.

Statistical analysis

Pooling allele frequencies (Peiris *et al.* 2011) were calculated for each SNP within each pool. Pooling allele frequency was used as an estimate of (or proxy for) the allele frequency that would be obtained by genotyping the population individually; however, the authors recognize that the relationship between pooling allele frequency and allele frequency is not exact. The R^2 was 0.95 or greater for 95 percent of the BovineHD beadarray SNPs in a pilot project in which pooling allele frequencies were compared to allele frequencies estimated from individual genotyping (J. W. Keele, unpublished data).

Mixed model methods were used to estimate differences in allele frequency between affected and unaffected cattle while accounting for technical variation, population stratification, binomial sampling and pool construction error (McDaneld *et al.* 2014), as follows:

$$y = X\mu + \varepsilon,$$

where μ is a vector of mean pooling allele frequency for cases and controls; y is a vector of pooling allele frequencies for each of n pools; X is an $n \times 2$ matrix of zeroes and ones indicating whether the pools are affected or unaffected; and ε is a vector of residuals distributed as MVN $(0, \sigma^2\boldsymbol{V})$, where MVN() is the multivariate normal distribution, σ^2 is a SNP-specific multiplier and $\boldsymbol{V}$ is the variance–covariance matrix characterizing the genetic similarities and differences among pools. $\boldsymbol{V}$ was estimated using cov() from R using pooling allele frequencies from all SNPs in the genome with a row for each SNP and a column for each pool. The F-statistic was computed as $= (y'(M - P)y(n - 2))/(y'Py)$, where $P = V^{-1} - V^{-1}X(X'V^{-1}X)X'V^{-1}$ and $M = V^{-1} - V^{-1}\mathbf{1}(\mathbf{1}'V^{-1}\mathbf{1})\mathbf{1}'V^{-1}$ and $\mathbf{1}$ is an $n \times 1$ vector with a value of 1 for every element. The P-value was calculated as the upper tail of the F distribution greater than or equal to the F-statistic with 1 numerator degree of freedom and $n - 2$ denominator degrees of freedom. The non-diagonal covariance matrix among pools ($\boldsymbol{V}$) controls for stratification effects at the individual SNP level; multipliers, σ^2, were estimated specifically for each SNP, so the scaling was adjusted to the distribution for each SNP. Statistical analysis was performed using R (Team R.D.C. 2011). Statistical thresholds for moderate ($P < 1 \times 10^{-5}$) and strong ($P < 5 \times 10^{-7}$) associations were established according to the Wellcome Trust Case Control Consortium (Consortium 2007). The nominal P-value required to achieve the 5 percent level genome-wide significance level was $P = 1.49 \times 10^{-7}$ based on an effective number of tests of 343 497, after adjusting for linkage disequilibrium using SIMPLEM (Gao *et al.* 2008). Quantile–quantile plots were generated using a script developed by Dr. Stephen Turner (http://gettinggeneticsdone.blogspot.com/2011/04/annotated-manhattan-plots-and-qq-plots.html). This script was developed for R (Team R.D.C. 2011). The estimate of genomic control (λ) was calculated to determine the degree of stratification in the analysis, according to Voorman *et al.* (2011).

Analysis of individually genotyped SNPs to establish whether the pooling procedure used was reliable was carried out using the procedure ALLELE from SAS (SAS Institute).

Based on the SNP association results, genes of interest in the genomic regions were identified within 100 kb from the position of the significant SNP using dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/). Genes of interest were queried in GWASdb (http://jjwanglab.org:8080/gwasdb/) to establish their potential role with diseases in human (Li *et al.* 2012).

Results and discussion

Results of the genome-wide association study are presented in Fig. 1. Genomic regions on chromosomes 1, 2, 6, 8, 10, 14, 15 and 18 were associated with presence or absence of persistent BVDV infection. All genomic regions had a moderate association ($P < 1 \times 10^{-5}$), except a SNP on chromosome 14, which had a strong association ($P < 5 \times 10^{-7}$) as well as a genome-wide significance at the 5 percent level ($P = 9.41 \times 10^{-8}$). The significant SNPs, their chromosome, position within the chromosome, the allele frequencies of the A allele in the affected and unaffected pools, their significance and the genes within 100 kb of the SNP are shown in Table 1.

A marker on chromosome 14 presented a strong association ($P = 9.41 \times 10^{-8}$) with persistent BVDV (Table 1).

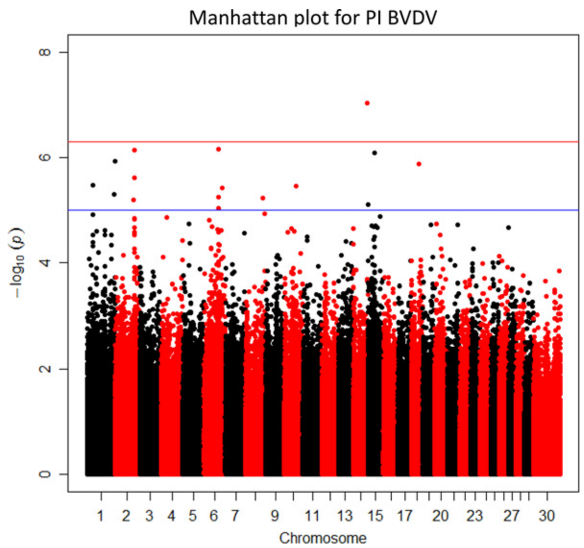

Figure 1 Manhattan plot of association of single nucleotide polymorphisms for persistent bovine viral diarrhea virus infection. Upper horizontal line represents a strong association threshold ($P = 5 \times 10^{-07}$ [$-\log10(P\text{-value}) = 6.3$]). Lower horizontal line represents a moderate association threshold ($P = 1 \times 10^{-5}$ [$-\log10(P\text{-value}) = 5.0$]). The X chromosome is represented with the number 30.

The marker resides within the *Raly RNA binding protein-like* (*RALYL*) gene sequence. *RALYL* is also known as the *heterogeneous nuclear ribonucleoprotein C (C1/C2)* (*HNRNPC*) gene. This gene is a member of the heterogeneous nuclear ribonucleoprotein family of genes (Jiang *et al.* 1998; Tenzer *et al.* 2013). It has been proposed that *HNRNPC* plays a major role in splicing and 3′ end formation of pre-mRNA (Jiang *et al.* 1998). Wu *et al.* (2013) indicated that *HNRNPC* is down-regulated in human host cells infected with H3N2 swine influenza virus. It is possible that BVDV have a similar effect on *HNRNPC*, and allele differences in the SNPs may contribute to differences in maintenance of persistent BVDV infections. The genomic region where this gene resides has not been associated with diseases in cattle other than a single report of a quantitative trait locus (QTL) for mastitis in dairy cattle (Schulman *et al.* 2004). The association of the SNP on chromosome 14 needs to be evaluated further in additional samples of persistent BVDV and controls to establish the association of sequence differences in the *RALYL* gene and persistent BVDV.

A region of chromosome 2 was identified with bovine respiratory disease and persistent BVDV (Neibergs *et al.* 2011). Neibergs *et al.* (2011), using microsatellite markers, identified a QTL associated with bovine respiratory disease on chromosome 2 in a large beef cattle half-sib family. The support interval of this QTL spanned the telomeric end of the chromosome. Neibergs *et al.* (2011) focused on the interval between 126.7 and 135.9 cM to assess the association of this genomic region with persistent BVDV. Using SNPs, Zanella *et al.* (2011) refined the location of this genomic region with persistent BVDV. Results from the present study support the findings by Neibergs *et al.* (2011) of a genomic region on chromosome 2 harboring a gene or group of genes associated with persistent BVDV.

Persistent BVDV produce a decrease in white blood cell counts in cattle (Zimmerman *et al.* 2006). Marker *rs109923296* resides within the *non-homologous end-joining factor 1* (*NHEJ1*) gene on chromosome 2 (Table 1). *NEHJ1* is one of the four genes responsible (along with the *DCLRE1C*, *LIG4* and *PRKDC*) for severe combined immunodeficiency (SCID) in humans (Cipe *et al.* 2014). DNA double-strand breaks are DNA lesions produced by cell sources, such as replication errors, or external damaging agents. DNA double-strand breaks are repaired by the non-homologous end-joining pathways. Failure to properly repair DNA double-strand breaks results in various forms of immunodeficiency (Buck *et al.* 2006). *NHEJ1* plays a major role in the non-homologous end-joining pathway (Cipe *et al.* 2014). Mutations in the *NHEJ1* gene have been associated with Cernunnos SCID (Buck *et al.* 2006). It has also been established that Cernunnos leucopenia, or reduced white blood cell counts, is associated with changes in the production of T cells (Vera *et al.* 2013). It has also been reported that T-cell numbers are reduced during the course of acute BVDV infection. Therefore, the *NHEJ1* gene is a likely candidate for further evaluation regarding impact of persistent infection on T-cell populations in persistently BVDV cattle.

Two regions on chromosome 6 were associated with persistent BVDV. Both regions are separated by approximately 2 mb. In one region, three SNPs residing near the *latrophilin 3* (*LPHN3*) gene were associated with persistent BVDV (Table 1). Latrophilin-3 is a member of the G protein-coupled receptors. Their function is to identify external cell molecules and initiate signal transduction (Matsushita *et al.* 1999). To date, this gene has not been associated with disease; however, mutations in *LPHN3* have been associated with lung adenocarcinoma in humans (Kan *et al.* 2010). Additional studies are necessary to establish whether this gene is associated with persistent BVDV. In another region of chromosome 6, a SNP residing in the *LIN54* gene was associated with persistent BVDV. The LIN core (LINC) is a complex that functions as a transcriptional repressor during cell cycle. *LIN54* is a core gene of LINC and has DNA binding activity (Schmit *et al.* 2009). Mutations on this gene have been associated with abnormal DNA binding activity and may have an effect on progression in cell cycle (Matsuo *et al.* 2012). It is possible that limiting progression of cell cycle can affect viral replication, which enhances the animal's ability to survive the persistent infection.

A marker residing in the *spermatogenesis-associated 5-like 1* (*SPATA5L1*) gene, on chromosome 10, was associated with persistent BVDV (Table 1). The *glycine amidinotransferase (L-arginine:glycine amidinotransferase)* (*GATM*) gene also resides in this genomic region. Kottgen *et al.* (2009) found an association of a SNP in the *GATM/SPATA5L1* locus with differences in glomerular filtration rate (GFR) estimated by serum creatinine in humans. GFR is a trait associated with chronic kidney disease. *GATM* encodes glycine amidinotransferase, an enzyme involved in creatine biosynthesis (Kottgen *et al.* 2009). Kottgen *et al.* (2009) indicated that, although this SNP is located in the *SPATA5L1* gene, it is likely associated with *GATM* due to strong linkage disequilibrium between both genes. Park *et al.* (2013), using a different SNP within the same genomic region, identified a similar association with GFR in an independent population. In chronic kidney disease in humans, RNA viruses from the *Flaviviridae* family trigger the release of interferons in the cell. Once the virus is in the cytoplasm, the virus's non-structural protein complex inactivates components of the toll-like receptor signaling complex, stopping the interferon antiviral mechanism, stalling the innate immune response. This blockage results in persistence of the virus beyond the acute phase, especially in animals with weakened or suppressed immune systems (Bruggeman 2007). The association of the SNP in the *GATM/SPATA5L1* locus could be evidence of this possibility.

One marker on chromosome 18 was associated with persistent BVDV (Table 1). The nearest gene (*KCTD15*) has been associated with obesity in humans (Paternoster *et al.* 2011). The association of this gene with persistent BVDV is unclear. Leucopenia has been detected in cattle affected with BVDV (Zimmerman *et al.* 2006). In a genome-wide association study, the *HYDIN* gene was proposed as a candidate for explaining variation for leucopenia in African American individuals (Reiner *et al.* 2011). This gene is in the genomic region where *KCTD15* resides on chromosome 18.

There were three SNPs associated with persistent BVDV on chromosome 1. There were no genes in the vicinity of one of the SNPs. Two SNPs at the telomeric end of chromosome 1 residing in two different genes were associated with persistent BVDV (Table 1). One is within the sequence of the *methyltransferase like 6* (*METTL6*) gene. There are few reports of the activity of *METTL6* associated with diseases. Tan *et al.* (2011) identified a SNP in this gene as associated with overall survival of patients with lung cancer. Tan *et al.* (2011) indicated that *METTL6* might influence DNA methylation and that methylation is related to better median overall survival of patients with lung cancer. A genetic tendency to survive may be what allows these animals to survive, while other persistently infected animals die as neonates. The second SNP in the telomeric region of chromosome 1 is located in the *TBC1 domain family, member 5* (*TBC1D5*) gene. This gene is a regulator of Rab GTPase activity. Seaman *et al.* (2009) indicated that *TBC1D5* inhibits the vacuolar protein sorting 35 homolog protein (VPS35/29/26), which selects cargo for endosome to Golgi retrieval. The role of *TBC1D5* in establishing a persistent infection needs to be further examined.

Two SNPs on chromosome 15 were associated with persistent BVDV. No genes were in the vicinity of one marker, whereas three genes were in the vicinity of the other (Table 1). Two of the three genes (*KCNC1* and *OTOG*) have been associated with auditive difficulties in humans, and the third gene (*MYOD1*) is associated with muscle regulation. It is unclear whether these genes have any association with persistent BVDV in cattle. A QTL for somatic cell score has been reported in this genomic region (Boichard *et al.* 2003). Additional studies are needed to establish the association of this genomic region with persistent BVDV.

There was one significant SNP on chromosome 8. There is no gene annotated within 100 Kb of this marker. A gene beyond 100 Kb is the *transmembrane protein 245* gene. This gene is part of the cellular signal transduction mechanism and could be associated with persistent BVDV. The SNP on chromosome 8 also neighbors the *toll-like receptor 4* (*TLR4*) gene. This gene is part of the toll-like receptor family of conserved glycoproteins that play a key role in the innate immune system. Their function is to sense the presence of pathogens and initiate the immune response (Turin & Riva 2008). It is possible that the association observed in the present study is related to suppression of this gene.

Results of SNPs by individual genotyping are shown in Table 2. Sixteen significant SNPs were identified in the genome-wide association analysis; however, eight provided useful information when individual genotyping was carried out. The other eight SNPs were either not selected by the Sequenom software to be analyzed or the genotyping was unsuccessful. Seven of the eight SNPs were significantly associated ($P < 0.05$) with persistent BVDV when a sample ($n = 384$) of the population was genotyped. The non-significant SNPs showed a tendency of significance ($P < 0.1$). Individual genotyping was carried out on 392 animals (196 affected and 196 unaffected). This is a sample (16.5%) of the population from the study. Therefore, *P*-values obtained with the sample should not be as significant as when the entire population is used. These results suggest that association of SNPs with persistent BVDV used in the population and the use of pooled samples is adequate to minimize costs. Given the results presented in Table 2, it was deemed unnecessary to pursue additional genotyping of other significant SNPs.

Allelic frequencies of the *A* allele were different for each SNP (Table 1). Estimated allele frequency differences for all significant SNPs were greater than 5 percent. A pool size of 100 animals minimizes cost while obtaining 80 percent statistical power to detect allele frequency differences of at least 5 percent (McDaneld *et al.* 2012). McDaneld *et al.* (2012) indicated that allele frequency of pools for affected and unaffected is biased relative to the actual allele

Table 2 Association of individual genotyping of single nucleotide polymorphisms (SNPs) in a sample of the population with persistent BVDV. Chromosome (BTA) and position, frequency of the minor allele in unaffected, and unaffected groups and significance are indicated for each SNP evaluated.

SNP	BTA	Position (bases)	Allele frequency (unaffected)	Allele frequency (affected)	*P*-value
Rs43226513	1	32138151	0.159	0.107	0.0526
Rs43287585	1	156065876	0.331	0.257	0.0365
Rs135524200	2	107608745	0.488	0.399	0.0262
Rs135783786	6	79369511	0.403	0.403	0.0006
Rs134088824	6	79385615	0.405	0.489	0.0052
Rs135088957	6	79389382	0.408	0.469	0.0015
Rs42255579	6	99531916	0.106	0.195	0.0014
Rs110572575	18	44427437	0.154	0.095	0.0209

frequencies, and the difference in pooling allele frequencies between affected and unaffected animals underestimates the true difference in allele frequencies. Therefore, pooling of samples was an efficient and conservative approach for a genome-wide association study.

The quantile–quantile plot for the genome-wide association study is shown in Fig. 2. Results on this plot indicate that no unrecognized population stratification or systematic differences in allele frequencies affected the analysis. To corroborate this observation, the assessment of genomic control for the quantile–quantile plot was calculated ($\lambda = 1.07$). When population stratification is absent, the estimate of genomic control should be similar to the standard test of independence (Devlin *et al.* 2001); this is $\lambda = 1$. Substantial inflation of statistics should be observed when model misspecifications or population stratification exists (Voorman *et al.* 2011). A genomic control adjustment should not considerably influence the results of the present study due to a relatively small inflation factor. To further investigate the potential for population stratification, we developed quantile–quantile plots separately for each chromosome (data not shown). The effect of population stratification, if it has an effect on the *P*-value distribution, should be similar for all chromosomes. Quantile–quantile plots by chromosome did not support the theory that population stratification affected the *P*-value distribution. An excess number of SNPs with small *P*-values occurred for only six of the 29 autosomal chromosomes. Chromosomes containing excess numbers of SNPs with small *P*-values were chromosomes 1, 2, 6, 13, 15 and 18. A better explanation for a small number of chromosomes with an excess of SNPs with small *P*-values would be SNPs in linkage disequilibrium with unobserved quantitative trait nucleotides affecting the trait. Therefore, pooling of samples was a feasible approach for a genome-wide association study for persistent BVDV.

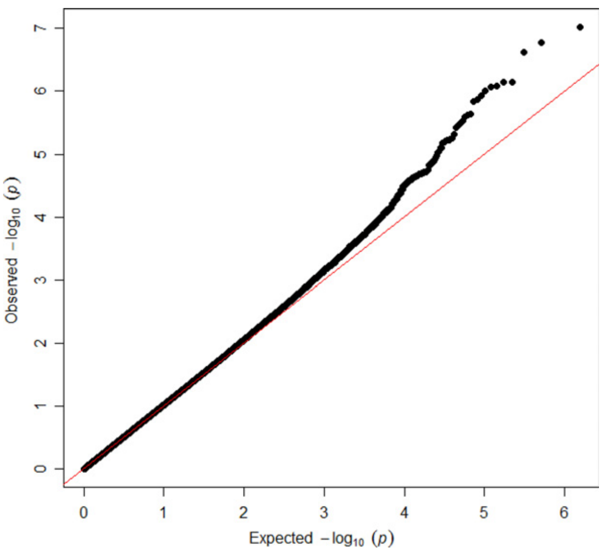

Figure 2 Quantile–quantile plot for association with persistent BVDV infection. The red line shows the expected distribution of *P*-values; the black line shows the observed distribution of *P*-values.

Statistically significant SNPs were associated between affected and unaffected animals with persistent BVDV. These markers reside within or in the vicinity of genes involved in several biological processes. Genes in which the significant SNPs reside are known to be involved in events that moderate immune responses, signal transduction, RNA splicing and DNA methylation. These processes may contribute to the animal's ability to survive the persistent infection.

Genomic regions identified in the present study are necessary to understand the disease in cattle. Additional samples could be collected to replicate the SNP associations observed in the present study. Current efforts at the National Animal Disease Center focus on developing animals with persistent BVDV under controlled conditions. Tissue from these animals will be used to establish differential expression of genes when compared with control animals. The relationship between the genes in the genomic regions identified in the present study and the differentially expressed genes will be established. Differential expression studies with acute high and low virulent BVDV strains have been conducted at the above-mentioned center. The relationship between the genes in the genomic regions of the present study and the differentially expressed genes will be ascertained. These comparisons could potentially provide an insight into how persistent BVDV affect cattle. These additional studies will confirm or reject the hypothesis that markers identified in the present study are associated with persistent BVDV in cattle.

Acknowledgement

The authors thank Randy Atchison from the NADC for technical assistance.

References

Blaschek M., Kaya A., Zwald N., Memili E. & Kirkpatrick B.W. (2011) A whole-genome association analysis of noncompensatory fertility in Holstein bulls. *J Dairy Sci* **94**, 4695–9.

Boichard D., Grohs C., Bourgeois F., Cerqueira F., Faugeras R., Neau A., Rupp R., Amigues Y., Boscher M.Y. & Leveziel H. (2003) Detection of genes influencing economic traits in three French dairy cattle breeds. *Genetics, Selection, Evolution* **35**, 77–101.

Bruggeman L.A. (2007) Viral subversion mechanisms in chronic kidney disease pathogenesis. *Clinical Journal of the American Society of Nephrology* **2**, S13–9.

Buck D., Malivert L., de Chasseval R. *et al.* (2006) Cernunnos, a novel nonhomologous end-joining factor, is mutated in human immunodeficiency with microcephaly. *Cell* **124**, 287–99.

Cipe F.E., Aydogmus C., Babayigit Hocaoglu A., Kilic M., Kaya G.D. & Yilmaz Gulec E. (2014) Cernunnos/XLF deficiency: a syndromic primary immunodeficiency. *Case Reports in Pediatrics* **2014**, 614238.

Consortium T.W.T. (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–83.

Devlin B., Roeder K. & Wasserman L. (2001) Genomic control, a new approach to genetic-based association studies. *Theoretical Population Biology* **60**, 155–66.

Finlay E.K., Berry D.P., Wickham B., Gormley E.P. & Bradley D.G. (2012) A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle. *PLoS One* **7**, e30545.

Fulton R.W., Hessman B., Johnson B.J., Ridpath J.F., Saliki J.T., Burge L.J., Sjeklocha D., Confer A.W., Funk R.A. & Payton M.E. (2006) Evaluation of diagnostic tests used for detection of bovine viral diarrhea virus and prevalence of subtypes 1a, 1b, and 2a in persistently infected cattle entering a feedlot. *Journal of the American Veterinary Medical Association* **228**, 578–84.

Gao X., Starmer J. & Martin E.R. (2008) A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology* **32**, 361–9.

Hessman B.E., Fulton R.W., Sjeklocha D.B., Murphy T.A., Ridpath J.F. & Payton M.E. (2009) Evaluation of economic effects and the health and performance of the general cattle population after exposure to cattle persistently infected with bovine viral diarrhea virus in a starter feedlot. *American Journal of Veterinary Research* **70**, 73–85.

Jiang W., Guo X. & Bhavanandan V.P. (1998) Four distinct regions in the auxiliary domain of heterogeneous nuclear ribonucleoprotein C-related proteins. *Biochimica et Biophysica Acta* **1399**, 229–33.

Kan Z., Jaiswal B.S., Stinson J. *et al.* (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. *Nature* **466**, 869–73.

Kottgen A., Glazer N.L., Dehghan A. *et al.* (2009) Multiple loci associated with indices of renal function and chronic kidney disease. *Nature Genetics* **41**, 712–7.

Li M.J., Wang P., Liu X., Lim E.L., Wang Z., Yeager M., Wong M.P., Sham P.C., Chanock S.J. & Wang J. (2012) GWASdb: a database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Research* **40**, D1047–54.

Matsuo T., Kuramoto H., Kumazaki T., Mitsui Y. & Takahashi T. (2012) *LIN54* harboring a mutation in CHC domain is localized to the cytoplasm and inhibits cell cycle progression. *Cell Cycle* **11**, 3227–36.

Matsushita H., Lelianova V.G. & Ushkaryov Y.A. (1999) The latrophilin family: multiply spliced G protein-coupled receptors with differential tissue distribution. *FEBS Letters* **443**, 348–52.

McClurkin A.W., Littledike E.T., Cutlip R.C., Frank G.H., Coria M.F. & Bolin S.R. (1984) Production of cattle immunotolerant to bovine viral diarrhea virus. *Canadian Journal of Comparative Medicine* **48**, 156–61.

McDaneld T.G., Kuehn L.A., Thomas M.G., Snelling W.M., Sonstegard T.S., Matukumalli L.K., Smith T.P., Pollak E.J. & Keele J.W. (2012) Y are you not pregnant: identification of Y chromosome segments in female cattle with decreased reproductive efficiency. *Journal of Animal Science* **90**, 2142–51.

McDaneld T.G., Kuehn L.A., Thomas M.G., Snelling W.M., Smith T.P., Pollak E.J., Cole J.B. & Keele J.W. (2014) Genome-wide association study of reproductive efficiency in female cattle. *Journal of Animal Science* published ahead of print March 26, 2014, doi:10.2527/jas.2013-6807.

Minozzi G., Williams J.L., Stella A., Strozzi F., Luini M., Settles M.L., Taylor J.F., Whitlock R.H., Zanella R. & Neibergs H.L. (2012) Meta-analysis of two genome-wide association studies of bovine paratuberculosis. *PLoS One* **7**, e32578.

Neibergs H., Zanella R., Casas E., Snowder G.D., Wenz J., Neibergs J.S. & Moore D. (2011) Loci on *Bos taurus* chromosome 2 and *Bos taurus* chromosome 26 are linked with bovine respiratory disease and associated with persistent infection of bovine viral diarrhea virus. *Journal of Animal Science* **89**, 907–15.

Park H., Kim H.J., Lee S., Yoo Y.J., Ju Y.S., Lee J.E., Cho S.I., Sung J., Kim J.I. & Seo J.S. (2013) A family-based association study after genome-wide linkage analysis identified two genetic loci for renal function in a Mongolian population. *Kidney International* **83**, 285–92.

Paternoster L., Evans D.M., Nohr E.A. *et al.* (2011) Genome-wide population-based association study of extremely overweight young adults–the GOYA study. *PLoS One* **6**, e24303.

Peiris B.L., Ralph J., Lamont S.J. & Dekkers J.C. (2011) Predicting allele frequencies in DNA pools using high density SNP genotyping data. *Animal Genetics* **42**, 113–6.

Reiner A.P., Lettre G., Nalls M.A. *et al.* (2011) Genome-wide association study of white blood cell count in 16,388 African Americans: the continental origins and genetic epidemiology network (COGENT). *PLoS Genetics* **7**, e1002108.

Schmit F., Cremer S. & Gaubatz S. (2009) LIN54 is an essential core subunit of the DREAM/LINC complex that binds to the cdc2 promoter in a sequence-specific manner. *FEBS Journal* **276**, 5703–16.

Schulman N.F., Viitala S.M., de Koning D.J., Virta J., Maki-Tanila A. & Vilkki J.H. (2004) Quantitative trait loci for health traits in Finnish Ayrshire cattle. *Journal of Dairy Science* **87**, 443–9.

Seaman M.N., Harbour M.E., Tattersall D., Read E. & Bright N. (2009) Membrane recruitment of the cargo-selective retromer subcomplex is catalysed by the small GTPase Rab7 and inhibited by the Rab-GAP TBC1D5. *Journal of Cell Science* **122**, 2371–82.

Stone R.T., Grosse W.M., Casas E., Smith T.P., Keele J.W. & Bennett G.L. (2002) Use of bovine EST data and human genomic sequences to map 100 gene-specific bovine markers. *Mammalian Genome* **13**, 211–5.

Tan X.L., Moyer A.M., Fridley B.L. *et al.* (2011) Genetic variation predicting cisplatin cytotoxicity associated with overall survival in lung cancer patients receiving platinum-based chemotherapy. *Clinical Cancer Research* **17**, 5801–11.

Team R.D.C. (2011) *R: A Language and Environment for Statistical Computing*. Team R.D.C., Vienna, Austria.

Tenzer S., Moro A., Kuharev J., Francis A.C., Vidalino L., Provenzani A. & Macchi P. (2013) Proteome-wide characterization of the RNA-binding protein RALY-interactome using the *in vivo*-biotinylation-pulldown-quant (iBioPQ) approach. *Journal of Proteome Research* **12**, 2869–84.

Turin L. & Riva F. (2008) Toll-like receptor family in domestic animal species. *Critical Reviews in Immunology* **28**, 513–38.

Vera G., Rivera-Munoz P., Abramowski V. *et al.* (2013) Cernunnos deficiency reduces thymocyte life span and alters the T cell repertoire in mice and humans. *Molecular and Cellular Biology* **33**, 701–11.

Voorman A., Lumley T., McKnight B. & Rice K. (2011) Behavior of QQ-plots and genomic control in studies of gene-environment interaction. *PLoS One* **6**, e19416.

Wu X., Wang S., Yu Y., Zhang J., Sun Z., Yan Y. & Zhou J. (2013) Subcellular proteomic analysis of human host cells infected with H3N2 swine influenza virus. *Proteomics* **13**, 3309–26.

Zanella R., Whitlock R.H. & Neibergs H.L. (2011) Short communication: refinement of genetic regions associated with *Mycobacterium avium* subspecies *paratuberculosis* tissue infection and tolerance to Johne's disease. *Journal of Dairy Science* **94**, 4230–6.

Zimmerman A.D., Boots R.E., Valli J.L. & Chase C.C. (2006) Evaluation of protection against virulent bovine viral diarrhea virus type 2 in calves that had maternal antibodies and were vaccinated with a modified-live vaccine. *Journal of the American Veterinary Medical Association* **228**, 1757–61.